AD-A009 866

A PREDICTION INTERVAL APPROACH TO DEFINING VARIABLES SAMPLING PLANS FOR FINITE LOTS REQUIRED TO BE OF HIGH QUALITY: SINGLE SAMPLING FOR GAUSSIAN PROCESSES

Kenneth W. Fertig, et al

Rocketdyne

Prepared for:

Office of Naval Research

1 February 1975

DISTRIBUTED BY:

NTIS

National Technical Information Service
U. S. DEPARTMENT OF COMMERCE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE	READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER / 2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER AD-A009 864
4. TITLE (and Subtitle) Prediction Interval Approach to Defining Variables Sampling Plans for Finite Lots Required to be of High Quality: Single Sampling for Gaussian Processes	5. TYPE OF REPORT & PERIOD COVERED Technical Report 2/1/75
	6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Kenneth W. Fertig, Nancy R. Mann	8. CONTRACT OR GRANT NUMBER(s) N00014-73-C-0474
9. PERFORMING ORGANIZATION NAME AND ADDRESS Rocketdyne, Division of Rockwell International, 6633 Canoga Avenue, Canoga Park, CA 91304	10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-042-321
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research, Department of the Navy, Arlington, Va 22217	12. REPORT DATE 2/1/75
	13. NUMBER OF PAGES 23
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)	15. SECURITY CLASS. (of this report) Unclassified
	15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Variables Sampling Plans
Finite Lots
Prediction Intervals
Gaussian Processes

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

See Title page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

A Prediction Interval Approach to Defining Variables Sampling Plans for Finite Lots Required to be of High Quality: Single Sampling for Gaussian Processes

by

Kenneth W. Fertig and Nancy R. Mann

ABSTRACT

A prediction-interval approach and an assumption of a Gaussian manufacturing process are used to derive a variables sampling plan applicable to finite lots required to be of very high quality. Values tabulated for calculating acceptance regions are such that, with high probability, accepted lots will have zero defects. Further, tables are given for selecting sample size for a specified combination of lot size and acceptable quality level. Comparisons show that, for a fixed risk level, substantial savings in required sample size can be effected over those specified by comparable hypergeometric sampling plans. Lots ranging in size from 5 to 100 are considered.

Bounds for the probability of accepting a lot with a fixed number of defects and coming from a prescribed manufacturing process are derived and prove to be very tight. Not only are these bounds useful in defining the required sample size in the sampling plan context, but the method of derivation has application to other areas of sampling, e.g. from truncated or stratified distributions.

The research presented herein was performed at Rocketdyne, a Division of Rockwell International, under the sponsorship of the Office of Naval Research, Contract number N00014-73-C-0474, Task NR-042-321.

1. INTRODUCTION

In the following, consideration is given to sampling from finite lots for which the underlying process distribution is Gaussian with unknown mean and variance. It is assumed that accepted lots are required to be of very high quality, that is, with high probability, accepted lots should have zero defects. An item will be said to be defective if its measured parameter, x, is less than some specified value L (greater than a specified U).

Suppose one were to make no attempt to use the knowledge concerning the distributional form of the process that generated the lot. Acceptance or rejection of the lot would then be based solely on the attributes data, i.e. whether or not the observed x's were less than L (or exceeded U). A required sample size would probably be determined for a specified combination of lot size, rejectable quality level and acceptable quality level by use of the hypergeometric distribution, perhaps from tables of Lieberman and Owen [8]. Use of binomial sampling plans which inherently make no provision for the finite size of the lot, can lead to sample-size requirements that are unnecessarily large, perhaps even exceeding the lot size.

Here it is assumed that we record the variables data, the sample measurements, $x_1, \ldots, x_n$, of the normal variate that determines whether or not an item is defective. One would expect that the required sample size associated with a specified combination of lot size and acceptable quality level based on the additional information $x_1, \ldots, x_n$ and an assumption of normality for the process distribution would be less than required solely on the basis of whether or not the x's are less than L (greater than U). What is desired, therefore, is a small-lot variables sampling plan (one that makes use of sample measured values and the distributional form of the process) that can be applied to lots required to be of very high quality. In the following, only a lower specification limit L will be discussed. The theory developed can be applied to upper specification limits by appropriate changes in notation and sign.

The small-lot Gaussian sampling plans of Fertig and Mann [3], the theory for which is described in [5], are based on a decision-theoretic approach and an economic loss function. Use of the loss function requires that rejected lots be screened and defective items be replaced by nondefectives. Because of this requirement, these sampling plans cannot be applied, for example, to situations of destructive testing. The sampling plans to be described in the sequel, however, do not depend on the utilization of a loss function and can be used even if the sampling situation involves destructive testing. Moreover, these plans are based on what will generally be considered a more classical approach involving none of the Bayesian methods required in the decision theoretic approach, e.g. the use of "optimal" priors on the process parameters to compute the posterior expectation of the number of defects remaining in the lot after sampling.

The new sampling plans are "Double Zero" sampling plans in the sense described by Ellis [2], that is, lots are required to have zero defects with high probability and detection of a defect in a sample requires rejection of the corresponding lot. Ellis [2] maintains that such plans, which are used currently at Pratt and Whitney Aircraft, "are especially effective in the metal working industry." The new plans differ from Ellis' in that these are based on actual variables data while his are not and here there is no consideration given to Ellis' "gray areas" of marginal quality.

2. BASIS OF THE SAMPLING PLANS

A sampling plan can be described in terms of its acceptance criteria and the manner of specifying the sample size as a function of lot size and given requirements on quality level. Here, the approach used in defining the acceptance criteria is based on the theory of classical prediction intervals. Thus, for a sample of size n from a lot of size N, the sample variables data $x_1, \ldots, x_n$, are used to predict whether or not the N-n unsampled items in the lot are all nondefective. The actual prediction interval is computed on the basis of the sampling distribution of

$$R_{n,N-n} \equiv (\overline{X} - Y_{1,N-n})/S \tag{2.1}$$

where $Y_{1,N-n}$ is the smallest observable value of the random variate X in the unsampled portion of the lot and $\overline{X}$ and S are the random variates with realizations $\overline{x}$ and s, respectively, calculated from $x_1, \ldots, x_n$. If we let $\theta_{\gamma,n,N-n}$ be the 100γth percentile of $R_{n,N-n}$, then $(\overline{X} - \theta_{\gamma,n,N-n}S, +\infty)$ is the 100γ percent lower prediction interval for $Y_{1,N-n}$.

Methodology for relatively easy and inexpensive computer calculation of $\theta_{\gamma,n,N-n}$ was developed by Fertig and Mann [6]. The procedure depends on the numerical integration of a single integral (shown in Section 5), in contrast to the earlier approach of Hahn [7] which involved a multivariate t-distribution and numerical integration of a double integral.

Care must be exercised in the interpretation of a prediction interval. Specifically, in this case it means that if pairs of samples of sizes n and N-n are created repeatedly by the manufacturing process, the 100 γ percent random prediction interval $(\overline{X} - \theta_{\gamma,n,N-n}S, +\infty)$ with $\overline{X}$ and S computed from the first sample of size n of the pair, will contain $Y_{1,N-n}$, the smallest observable variate in the corresponding second sample of the pair, 100γ percent of the time. If the prediction interval contains the smallest observation $y_{1,N-n}$, then of course, it contains all the observations in the second sample, the unsampled portion of the lot in the present context.

2.1 Acceptance Criteria

For the sampling-plan tables presented in this paper, we have set $\gamma = 0.90$. The value of $\theta_{\gamma,n,N-n}$ for each n and N considered is given in Table 1. The basis for choosing the sample size n will be discussed in section 2.2. Once the sample $x_1, \ldots, x_n$ has been observed and $\bar{x}$ and s computed, the following specific acceptance criteria for the lot can be applied.

Accept the lot if

$$X_i > L \quad \text{for } i=1, \ldots, n \tag{2.2}$$

and

$$\bar{X} - \theta_{\gamma,n,N-n}\, s > L. \tag{2.3}$$

These criteria follow directly from the consideration that only those lots that have no defective items in them are acceptable. The first condition (2.2) specifies that there must be no defectives in the sample and the second condition specifies that the sample must predict with a high degree of confidence (90 percent) that there are no defectives in the remaining portion of the lot.

2.2 Selection of Sample Size

In the classical operating-characteristics approach to indexing sampling plans, the acceptance criteria and sample size are chosen so that the producer's

TABLE 1. FINITE LOT VARIABLES SAMPLING PLAN ACCEPTANCE FACTORS, θ_γ, n, N-n, FOR γ = 0.90 AND ACCEPTANCE QUALITY LEVEL PERCENTILES, AQLP, FOR 1 - α = 0.95

Lot Size	Sample Size	θ (0.90)	AQLP (0.95)
5	2	7.594	14.949
	3	3.066	5.460
	4	1.831	3.156
6	2	8.730	17.166
	3	3.615	6.387
	4	2.484	4.183
	5	1.680	2.727
7	2	9.619	18.905
	3	4.009	7.058
	4	2.873	4.789
	5	2.240	3.621
8	2	10.345	20.323
	3	4.316	7.580
	4	3.150	5.223
	5	2.567	4.110
	6	2.105	3.286
9	2	10.953	21.513
	3	4.565	8.006
	4	3.364	5.560
	5	2.798	4.454
	6	2.399	3.722
	7	2.020	3.057
10	2	11.476	22.535
	3	4.775	8.364
	4	3.538	5.834
	5	2.975	4.720
	6	2.606	4.021
	7	2.294	3.463
11	2	11.932	23.426
	3	4.954	8.672
	4	3.684	6.065
	5	3.120	4.936
	6	2.764	4.251
	7	2.485	3.738
	8	2.221	3.274
12	2	12.335	24.216
	3	5.112	8.941
	4	3.809	6.264
	5	3.241	5.118
	6	2.893	4.436
	7	2.631	3.916
	8	2.402	3.533
	9	2.167	3.131
13	2	12.696	24.922
	3	5.252	9.180
	4	3.919	6.438
	5	3.313	5.274
	6	3.001	4.593
	7	2.750	4.113
	8	2.540	3.728
	9	2.341	3.377
14	2	13.022	25.561
	3	5.377	9.395
	4	4.016	6.591
	5	3.436	5.112
	6	3.093	4.726
	7	2.849	4.351
	8	2.652	3.884
	9	2.473	3.562
	10	2.294	3.254

Lot Size	Sample Size	θ (0.90)	AQLP (0.95)
15	2	13.320	26.143
	3	5.491	9.590
	4	4.104	6.732
	5	3.517	5.534
	6	3.174	4.844
	7	2.935	4.374
	8	2.745	4.014
	9	2.580	3.711
	10	2.422	3.432
	11	2.257	3.155
16	2	13.592	26.676
	3	5.594	9.768
	4	4.184	6.859
	5	3.590	5.644
	6	3.246	4.949
	7	3.010	4.480
	8	2.826	4.126
	9	2.669	3.834
	10	2.525	3.574
	11	2.382	3.326
17	2	13.844	27.169
	3	5.689	9.951
	4	4.256	6.974
	5	3.656	5.743
	6	3.311	5.043
	7	3.076	4.573
	8	2.896	4.223
	9	2.746	3.939
	10	2.611	3.691
	11	2.482	3.463
	12	2.349	3.238
18	2	14.077	27.626
	3	5.778	10.083
	4	4.323	7.081
	5	3.716	5.835
	6	3.370	5.129
	7	3.136	4.658
	8	2.959	4.310
	9	2.813	4.031
	10	2.685	3.792
	11	2.565	3.576
	12	2.442	3.371
	13	2.322	3.164
19	2	14.295	28.051
	3	5.859	10.223
	4	4.385	7.179
	5	3.772	5.919
	6	3.424	5.207
	7	3.190	4.735
	8	3.015	4.388
	9	2.873	4.112
	10	2.750	3.880
	11	2.637	3.673
	12	2.528	3.481
	13	2.418	3.293
20	2	14.498	28.449
	3	5.936	10.355
	4	4.445	[illegible]
	5	3.823	5.997
	6	[illegible]	[illegible]
	7	3.239	4.805
	8	3.066	4.458
	9	2.926	4.185
	10	2.807	3.957
	11	[illegible]	[illegible]
	12	[illegible]	[illegible]
	13	2.197	3.400
	14	[illegible]	3.226

Lot Size	Sample Size	θ (0.90)	AQLP (0.95)
25	2	15.353	30.123
	3	6.256	10.905
	4	4.682	7.652
	5	4.035	6.317
	6	3.675	5.573
	7	3.438	5.087
	8	3.267	4.737
	9	3.134	4.468
	10	3.025	4.250
	11	2.931	4.066
	12	2.847	3.905
	13	2.770	3.761
	14	2.696	3.627
	15	2.625	3.501
	16	2.553	3.378
	17	2.478	3.255
30	2	16.019	31.427
	3	6.504	11.331
	4	4.866	7.945
	5	4.196	6.562
	6	3.825	5.794
	7	3.585	5.295
	8	3.413	4.940
	9	3.281	4.669
	10	3.175	4.452
	11	3.086	4.273
	12	3.009	4.119
	13	2.940	3.984
	14	2.877	3.863
	15	2.818	3.752
	16	2.762	3.649
	17	2.707	3.551
	18	2.652	3.456
	19	2.597	3.363
	20	2.540	3.270
	21	2.481	3.175
35	2	16.562	32.490
	3	6.706	11.679
	4	5.015	8.183
	5	4.326	6.758
	6	3.946	5.970
	7	3.701	5.460
	8	3.527	5.098
	9	3.395	4.824
	10	3.290	4.606
	11	3.203	4.427
	12	3.129	4.277
	13	3.064	4.146
	14	3.005	4.029
	15	2.952	3.925
	16	2.902	3.829
	17	2.855	3.741
	18	2.810	3.657
	19	2.766	3.578
	20	2.722	3.502
	21	2.678	3.427
	22	2.634	3.353
	23	2.589	3.279
	24	2.541	3.204

TABLE 1. (Concluded)

Lot Size	Sample Size	θ (0.90)	AQLP (0.95)
40	2	17.019	33.386
	3	6.876	11.971
	4	5.139	8.382
	5	4.433	6.922
	6	4.045	6.116
	7	3.796	5.596
	8	3.620	5.228
	9	3.487	4.950
	10	3.382	4.731
	11	3.296	4.551
	12	3.223	4.401
	13	3.160	4.271
	14	3.104	4.158
	15	3.054	4.056
	16	3.008	3.965
	17	2.964	3.880
	18	2.924	3.802
	19	2.885	3.729
	20	2.848	3.660
	21	2.811	3.594
	22	2.775	3.530
	23	2.739	3.468
	24	2.703	3.407
	25	2.667	3.346
	26	2.629	3.286
	27	2.590	3.224
45	2	17.413	34.156
	3	7.02[illegible]	12.223
	4	5.246	8.553
	5	4.526	7.062
	6	4.130	6.240
	7	3.877	5.711
	8	3.699	5.337
	9	3.565	5.056
	10	3.460	4.835
	11	3.373	4.654
	12	3.301	4.503
	13	3.239	4.374
	14	3.185	4.261
	15	3.136	4.161
	16	3.092	4.072
	17	3.051	3.990
	18	3.013	3.915
	19	2.978	3.846
	20	2.944	3.781
	21	2.911	3.719
	22	2.879	3.660
	23	2.848	3.604
	24	2.818	3.550
	25	2.787	3.497
	26	2.757	3.445
	27	2.726	3.394
	28	2.695	3.343
	29	2.663	3.291
	30	2.630	3.239
	31	2.596	3.187

Lot Size	Sample Size	θ (0.90)	AQLP (0.95)
50	2	17.757	34.830
	3	7.150	12.443
	4	5.340	8.702
	5	4.606	7.184
	6	4.204	6.348
	7	3.947	5.811
	8	3.767	5.432
	9	3.632	5.147
	10	3.526	4.924
	11	3.440	4.742
	12	3.368	4.590
	13	3.306	4.461
	14	3.252	4.348
	15	3.204	4.249
	16	3.161	4.161
	17	3.122	4.080
	18	3.086	4.007
	19	3.052	3.940
	20	3.020	3.877
	21	2.990	3.818
	22	2.961	3.763
	23	2.933	3.710
	24	2.906	3.660
	25	2.879	3.611
	26	2.853	3.564
	27	2.827	3.518
	28	2.801	3.473
	29	2.774	3.428
	30	2.748	3.384
	31	2.721	3.340
	32	2.693	3.296
	33	2.665	3.252
	34	2.635	3.207
60	2	18.336	35.964
	3	7.366	12.814
	4	5.497	8.954
	5	4.741	7.390
	6	4.328	6.530
	7	4.061	5.978
	8	3.880	5.590
	9	3.712	5.299
	10	3.633	5.071
	11	3.548	4.886
	12	3.476	4.733
	13	3.414	4.602
	14	3.361	4.489
	15	3.315	4.391
	16	3.273	4.303
	17	3.235	4.221
	18	3.201	4.152
	19	3.169	4.087
	20	3.140	4.026
	21	3.112	3.970
	22	3.086	3.918
	23	3.061	3.869
	24	3.038	3.822
	25	3.015	3.778
	26	2.993	3.736
	27	2.971	3.695
	28	2.950	3.656
	29	2.930	3.618

Lot Size	Sample Size	θ (0.90)	AQLP (0.95)
60	30	2.909	3.581
	31	2.889	3.545
	32	2.869	3.511
	33	2.849	3.476
	34	2.828	3.442
	35	2.808	3.408
	36	2.787	3.374
	37	2.766	3.340
	38	2.744	3.306
	39	2.722	3.272
	40	2.699	3.237
	41	2.676	3.202
70	2	18.814	36.901
	3	7.544	13.120
	4	5.627	9.161
	5	4.852	7.558
	6	4.429	6.678
	7	4.160	6.114
	8	3.972	5.718
	9	3.832	5.421
	10	3.722	5.189
	11	3.635	5.002
	12	3.562	4.846
	13	3.501	4.715
	14	3.448	4.601
	15	3.402	4.502
	16	3.360	4.414
	17	3.323	4.335
	18	3.290	4.264
	19	3.260	4.200
	20	3.231	4.140
	21	3.204	4.085
	22	3.180	4.034
	23	3.157	3.987
	24	3.135	3.942
	25	3.114	3.900
	26	3.094	3.860
	27	3.075	3.822
	28	3.056	3.785
	29	3.038	3.750
	30	3.021	3.717
	31	3.003	3.684
	32	2.986	3.653
	33	2.970	3.622
	34	2.953	3.593
	35	2.937	3.563
	36	2.92[illegible]	3.535
	37	2.904	3.507
	38	2.888	3.479
	39	2.871	3.451
	40	2.855	3.424
	41	2.838	3.396
	42	2.821	3.370
	43	2.804	3.342
	44	2.786	3.315
	45	2.768	3.287
	46	2.750	3.260
	47	2.731	3.231

and consumer's risks satisfy certain constraints. Specifically, the producer's risk, the probability of rejecting lots at the acceptable quality level (AQL) must be less than or equal to α and the consumer's risk, the probability of accepting lots at the rejectable quality level (RQL) must be less than or equal to $1-\gamma$. For each sample size, the acceptance region given in (2.1) is designed to give the consumer lots of high quality. It is granted that (2.2) and (2.3) do not lead to the classical protection criteria at the RQL mentioned above, but they do lend themselves to a consumer's-risk interpretation. Specifically, only those lots that the user predicts not to be of rejectable quality (one or more defects) are accepted. Here, the consumer is "risking" a faulty prediction at the $\underline{100\gamma}$ th confidence level.

Taking this as a proper consumer's risk control, we are free to choose the sample size to control the producer's risk. Here we can closely follow the classical interpretation, in that we require that lots with zero defects will be accepted at least 100 $(1-\alpha)$ percent of the time. In order to compute this probability, however, we must specify the quality of the process the lot comes from as measured by p, the average fraction defective. Letting $\Phi(\cdot)$ be the cummulative standard normal distribution, μ and σ the mean and standard deviation of the manufacturing process, we have that $p=\Phi\ ((L-\mu)/\sigma)$. If we let k be the number of defects in the lot, we have that the sample size, n, must be such that:

$$P\ \{\text{accepting lot}\ |\ k=0,\ p\} \geq 1-\alpha \qquad (2.4)$$

If the number of defects in the lot were not specified as a condition in (2.4) and only (2.3) were used as an acceptance criterion, then the probability of accepting the lot could be computed directly as a function of p using the non central t-distribution. However, this is not the case. Furthermore, an exact expression for the left hand side of (2.4) is not available in general. Fortunately, upper and lower bounds for the left hand side of (2.4) are derivable. These bounds prove to be quite tight for all values of n and N of practical importance in the sampling-plan application. A theorem giving expressions for the bounds is proved in Section 4 for general k. If we let $\underline{P}$ be the lower bound and require that n be such that

$$\underline{P} \geq 1-\alpha\ , \tag{2.5}$$

then (2.4) is satisfied. This is exactly what has been done to generate the values in Table 1. Specifically, Table 1 gives to each N and n tabulated, the acceptance parameter $\theta_{\gamma,n,N-n}$ for $\gamma=0.90$ and the acceptable quality level percentile (AQLP) defined as

$$z_p = (L-\mu)/\sigma \tag{2.6}$$

with $p = \Phi(z_p)$ for $1-\alpha = 0.95$.

It is interesting to note that the upper bound $\overline{P}$ for the probability of acceptance is, for many combinations of N and n shown in Table 1, equal to $\underline{P}$. It is for these combinations of values of n and N that (2.4) is an exact equality. This occurs for all n less than N/2 up to N=25, for all n less than N/3 up to N=50, for all n less than N/4 up to N=80 and for all $n \leq 22$ for N=90 and 100. In all cases the difference is less than 1% between $\overline{P}$ and $\underline{P}$ for $n \leq N/2$, and less than 0.08% for $n \leq N/3$. Values for $1-\alpha=0.90$ as well as all values of $\overline{P}$ are given in Fertig and Mann [5].

Before we proceed with a proof of the theorem referred to above, we give an example of use of the tables.

3. EXAMPLE OF USE OF THE TABLES

Consider a lot of size 50. Suppose one wishes to reject the lot even if it has only one defect. The binomial plan, which ignores the finite size of the lot, would require a sample of size 114 in order to reject a lot ninety percent of the time that is 2%=100 (1/50) % defective. The hypergeometric plan would require a sample of size 45 to insure a 90% rejection rate. Since it is required to have such high quality in the accepted lot, one might wish to use a plan based on the assumption that the manufacturing process producing the lot is very good. If variables data are being recorded, and it can be assumed that these data are normally distributed, then the procedure presented herein will apply. Specifically, referring to Table 1, we find that a sample size of 9, for example, will guarantee a 95% acceptance rate if the manufacturing process mean is at least 5.147 σ above the lower specification limit. We would accept the lot if the corresponding sample, $x_1, \ldots, x_9$, had no defects and, from Table 1, its mean was at least 3.632 sample standard deviations above L. From this latter criterion we see why the process mean must be better than 5σ above L in order to guarantee a 95% acceptance rate.

At this point, it might be appropriate to discuss a criticism of the above procedure, namely expecting the normality assumption to hold at the extreme tails of the process distribution (5 σ from the mean). In fact, however, normality in the tails is not as necessary as one might first think to maintain accuracy in the calculation of the probability of acceptance. The normality assumption in the tails is used only in converting from the AQLP, $(\mu - L)/\sigma$, to the process fraction defective, p. In fact, since p will tend to be very small, it can be off by many orders of magnitude because of a lack of normality in the tails and still the probability of acceptance will remain relatively unaffected. This can be seen by inspecting the binomial probabilities in (4.4) and (4.5). The important parameter in the calculation of the probability of acceptance is not the fraction defective per se, but rather the acceptable quality level percentile $(L-\mu)/\sigma$ which is directly proportional to the noncentrality parameter in (4.4). It is this parameter which must be controlled by the manufacturer as well as the process average defective in order to guarantee a high probability of acceptance. It is for this reason that the tables are given in terms of AQLP rather than the AQL. Even if the quality is not as good as the AQLP, all that occurs is a higher rejection rate, essentially what is occurring with the hypergeometric plan wherein practically the entire lot is being screened. On the other hand, if the process quality is as good as the AQLP, a substantial savings in sample size has been effected.

Finally, we should remark that requiring very high manufacturing quality levels is not unreasonable in most cases in which lots would be unacceptable if they contained even a single defect.

4. THE BOUNDS ON THE PROBABILITY OF ACCEPTANCE OF A LOT

In this section we prove the following theorem which gives expressions for bounds on the probability of acceptance of a lot of size N on the basis of (2.2) and (2.3) when we specify fraction defective p in the process, number of defectives k in the lot and confidence level γ that an accepted lot has zero defectives. We note that this theorem may well be of interest outside the sampling-plan context. It applies generally to truncated-sampling applications and stratified-sampling problems.

Theorem

Let $T\,(t|v,\delta)$ be the cumulative non-central t-distribution with v degrees of freedom and noncentrality parameter δ. Let z_p be the 100p th percentile of the standard normal distribution. Let θ be the variables acceptance factor defined by (2.3) for a lot of size N and sample of size n. Consider lots that have k defects and come from a process which produces fraction p defective. The probability of accepting such lots satisfies:

$$\underline{P} \leq P\,\{\text{accepting lot } |k,p\} \leq \overline{P} \tag{4.1}$$

where

$$\overline{P} = \min\,(1,P^*) \tag{4.2}$$

$$\underline{P} = P^*-D, \tag{4.3}$$

with

$$P^* = \frac{H(0;k,n,N)}{B(0;p,n)}\left[1-T(\sqrt{n}\,\Theta|n-1,-\sqrt{n}z_p)\right] \tag{4.4}$$

and

$$D = \frac{H(0;k,n,N)}{B(0;p,n)}\sum_{\ell=1}^{m}\frac{\left[1-T(\sqrt{n-\ell}\,\Theta_\ell^*|n-\ell-1,-\sqrt{n-\ell}\,z_p)\right]\,B(\ell;p,n)}{B(0;\,p,n-\ell)}, \tag{4.5}$$

and

$$H(\ell;k,n,N) = \frac{\binom{k}{\ell}\binom{N-k}{n-\ell}}{\binom{N}{n}}\quad, \tag{4.6}$$

$$B(\ell;p,n) = \binom{n}{\ell} p^{\ell} (1-p)^{n-\ell}, \tag{4.7}$$

$$\Theta_{\ell}^{*} = \left(n\Theta \frac{(n-\ell-1)/(n-\ell)}{(n-\ell)(n-1)-n\ell\Theta^2}\right)^{1/2} \tag{4.8}$$

$$m = \max \left\{\text{all intergers} < \frac{n(n-1)}{n-1+n\Theta^2}\right\} \quad . \tag{4.9}$$

The values of Θ in (4.4) and of Θ_{ℓ}^{*} in (4.5), as can be seen from (4.1), depend upon the specified value of the confidence level γ as well as n and N-n. Also, as will be seen subsequently, summation on ℓ in (4.5) should be taken as zero if m=0. For this case $\underline{P}=\overline{P}$ and, therefore, the probability of accepting the lot can be computed exactly.

To prove (4.1), we first note that since the events $\ell=0$ and $\ell\neq0$ are disjoint all-inclusive events, then

$$\begin{aligned} &P\ \{\text{accepting lot } |k,p\} &(4.10)\\ &\quad = P\ \{\text{accepting lot } |\ell=0,k,p\}\ P\ \{\ell=0|k,p\}\\ &\quad + P\ \{\text{accepting lot } |\ell\neq0,k,p\}\ P\ \{\ell\neq0|k,p\}\end{aligned}$$

Since we reject the lot if $\ell \geq 1$, the second term on the right in (4.10) is zero. When k is given, the probability distribution of ℓ is hypergeometric. Therefore,

$$P\ \{\ell=0\ |k,p\} = \frac{\binom{k}{\ell}\binom{N-n}{k-\ell}}{\binom{N}{n}}$$

$$\equiv H(0;\ k,n,N)$$

When $\ell=0$, the decision whether or not to accept the lot rests on the value of the random variable $L^* = \overline{X}\text{-OS}$. Thus, we see that

$$P\ \{\text{accepting lot}\ |k,p\} \tag{4.11}$$
$$= P\ \{L^*>L|\ell=0,k,p\}\ H(0;k,n,N)$$

Finding the bounds in (4.1) now reduces to finding bounds for the probability that L^* is greater than L, given that it is based on a sample with no defects coming from a lot with k defects produced by a process with p fraction defective. Since we assume the items in the lot are stochastically independent observations from the process and since we are considering a series of lots all from the same process, sampling from the lot and requiring the sample to have zero defects is the same as sampling from the process and still requiring the sample to have zero defects. That is

$$P\ \{L^*>L|\ell=0,\ k,\ p\} = P\ \{L^*>L|\ell=0,\ p\}$$

where the expression on the right can be computed assuming a sample of size n is taken from a process of quality P with the restriction that no defects are in the sample. But since the events $\underset{\sim}{\ell}=\ell$ for $\ell=0,..,n$ are disjoint all-inclusive event for this procedure, we see that

$$P\ \{L^*>L|\ell=0,p\} = \frac{P\ \{L^*>L|p\}}{P\ \{\underset{\sim}{\ell}=0|p\}} \tag{4.12}$$
$$-\sum_{\ell=1}^{n} \frac{P\{L^*>L|\underset{\sim}{\ell}=\ell,p\}\ P\{\underset{\sim}{\ell}=\ell|p\}}{P\ \{\underset{\sim}{\ell}=0|p\}}$$

In order to proceed, we need the following three lemmas.

Lemma 1 Consider a sample of size n_1 from a process of quality p. Let $L_1^* = \overline{X}_1-\Theta S_1$ where $\overline{X}_1$ and S_1 are the observable sample mean and standard deviation $\left(\overline{X}_1=\sum_L X_i/n_1,\ S_1^2=\sum_L(X_i-\overline{X}_1)^2/(n_1-u)\right)$,

Then

$$P\ \{L_1^*>L|p\} = 1-T\ (\sqrt{n_1}\ \Theta|n_1-1,\cdot\ \sqrt{n_1}\ z_p)$$

Lemma 2 $P\{L_1^* > L | p\} < P\{L_1^* > L | \ell_1 = 0, P\}$

where ℓ_1 is the number of defects in the sample of size n_1

Lemma 3 $P\{L_1^* > L | \underset{\sim}{\ell}_1 = \ell_1, p\} < P\{L_1^{*\prime} > L | x_1, \ldots, x_{\ell_1} = L, \ell_1^{\prime} = 0, p\}$

where $L_1^{*\prime}$ is formed by replacing all observations in the sample that are less than L with the value L itself. (Thus, $\ell_1^{\prime}$, the number of defects remaining, is zero.)

The latter two lemmas follow easily if one notes that L_1^* is an increasing function of each of the sample observations. Thus restricting these observations to be above certain limits increases the probability of observing an L_1^* greater than L.

The first lemma follows from the definition of a non-central t-variate. Specifically,

$$P\{L_1^* > L | p\} = P\{\overline{X}_1 - \theta s_1 > L | p\}$$

$$= P\left(\frac{\sqrt{n}_1 \frac{\overline{X}_1 - \mu}{\sigma} - \sqrt{n}_1 \frac{L-\mu}{\sigma}}{s_1/\sigma} > \sqrt{n}_1 \theta \right)$$

Since $\frac{L-\mu}{\sigma} = z_p$, $\sqrt{n}(\overline{X}-\mu)/\sigma$ is $N(0,1)$, and s_1/σ is the square root of an independent chi square variate divided by its degrees of freedom $(n_1 - 1)$, the lemma follows directly by recalling that $t = (z+\delta)/x_\nu^2/\nu$ is a non-central t variate if z is standard normal and χ_ν^2 is an independent chi square with degrees of freedom ν.

Since the distribution of ℓ, given the sample comes from a process of quality p, is binomial with parameters p and n, we see from Lemma 1 that (4.12) implies

$$P\{L^*>L|\ell=0,p\} \le (1-T(\sqrt{n}\theta|n-1,-\sqrt{n}z_p))/B(0;n,p) \tag{4.13}$$

From (4.11) and (4.13) we see we can bound the probability of accepting the lot from above by the expression for P^* given in (4.4). This establishes (4.2).

In order to establish a lower bound or the probability of accepting the lot, one could use Lemma 2 directly in (4.11). A sharper bound, however, will in general be given by considering the summation in (4.12). Let L^* be based on an observable sample with ℓ defectives, $X_1,..X_\ell < L$ and $n-\ell$ nondefectives $X_{\ell+1},..., X_n \ge L$. Let $\bar{X}_2 = \sum_{i\ge\ell+1} X_i/(n-\ell)$, $S_2^2 = \frac{1}{(n-\ell-1)} \sum_{i\ge\ell+1}(X_i-\bar{X}_2)^2$.

We now seek to compute $P\{L^{*\prime}>L|X_1,...,X_\ell=L,\ell'=0,p\}$ defined in Lemma 3. After some algebra, we find that using Lemma 3 gives

$$P\{\bar{X}-\theta S>L|\underset{\sim}{\ell}=\ell,p\}$$

$$\le P\left\{\left(\frac{n-\ell}{n} - \frac{\theta^2\ell}{n-1}\right)(\bar{X}_2-L)^2 > \frac{n-\ell-1}{n-\ell}\,\frac{n}{n-1}\,\theta^2\, S_2^2 \mid \ell_2=0,p\right\}$$

This probability can be non-zero only if $\frac{n-\ell}{n} - \frac{\theta^2\ell}{n-1} > 0$. Thus, ℓ must be strictly less than $n(n-1)/(n-1+n\theta^2)$. This is the value given for m in (4.9). Since the number of defects, ℓ_2, in the portion of the sample to be used to compute $\bar{X}_2$ is zero, we have $\bar{X}_2 \ge L$. Therefore, we see that for $\ell \le m$,

$$P\{\bar{X}-\theta S>L|\underset{\sim}{\ell}=\ell,p\}$$

$$\le P\left\{(\bar{X}_2-L)/S_2 > n\theta\, \frac{(n-\ell-1)/(n-\ell)}{(n-\ell)(n-1)-n\ell\theta^2}^{1/2} \right.$$

$$\left. |\ell_2 = 0,p\right\}$$

Since $\bar{X}_2$ and S_2 are based on an observable sample of size $n_2 = n-\ell$ that has zero defects, we can apply Lemma 2 and a relation similar to the first half of this proof (in which the expression for P^* is derived) to find that

$$P\ \{\overline{X}\text{-}\Theta S>L\,|\,\underset{\sim}{\ell}=\ell,p\} \leq \frac{P\{\overline{X}_2-\Theta_\ell^* S_2>L\,|\,p\}}{P\{\ell_2=0\,|\,p\}}$$

Here, the probability in the numerator on the right is now computed with no restriction on the number of defects in the sample of size $n_2=n-\ell$. Therefore, we can apply Lemma 1 to get for each $\ell=1,\ldots,m$

$$P\ \{X\text{-}\Theta S>L\,|\,\underset{\sim}{\ell}=\ell,p\} \leq \frac{(1-T(\sqrt{n-\ell}\,\Theta_\ell^*\,|\,n-\ell-1,-\ \sqrt{n-\ell}\,z_p))}{B(0;p,n-\ell)} \tag{4.14}$$

and

$$P\ \{X\text{-}\Theta S>L\,|\,\underset{\sim}{\ell}\ \ell,p\} = 0 \tag{4.15}$$

for $\ell= m+1,\ \ldots,\ n.$

From (4.15) and the definition of m, we see that the sum in (4.14) should be taken as zero if m=0. For $m\geq 1$, we see from (4.11), (4.12) (4.14), and (4.15), that D is as given in (4.5). This completes the proof of the theorem.

5. CALCULATION OF TABLE ENTRIES

The distribution of the random variate $R_{n,N-n} = (\bar{X}-Y_{1,N-n})/S$, given by (2.1), which provides one of the criteria for acceptance or rejection of a lot of size N on the basis of a sample of size n, is considered by Fertig and Mann [6]. They show that

$$P(R_{n,N-n}<\Theta) =$$

$$\frac{N-n}{\sqrt{2\pi}} \int_{-\infty}^{\infty} [1-\Phi(z)]^{N-n-1} \exp(-z^2/2) T(\sqrt{N}\,\Theta | n-1, -\sqrt{n}\,z) dz, \quad (4.1)$$

where $T(\cdot|\nu,\delta)$ indicates the cumulative noncentral t-distribution function with degrees of freedom and noncentrality parameter δ and $\Phi(\cdot)$ indicates the cumulative standard normal distribution function with mean zero and variance one.

For each combination of N and n given in Table 1 $\underline{P}(R_{n,N-n}<\Theta) = .90$ was solved iteratively for Θ by use of Muller's [10] method, a multipoint iteration scheme which determines a next guess by using inverse parabolic interpolation based on the previous two iterates and their midpoint. A first guess for the value of Θ, was provided by the approximation suggested in Mann, Schafer and Singpurwalla [9] and investigated by Fertig and Mann [6]. Sixty-four point Hermite Gauss quadrature was used to evaluate the integral given by equation (4.1), the function $\Phi(\cdot)$ having been determined from $\mathrm{erf}(x) = 2\Phi(x\sqrt{2})-1$, $x \geq 0$, with the error function erf $(\cdot)$ computed using a double precision IBM subroutine and the noncentral t-distribution function computed using an algorithm of B.E. Cooper [1]. Where the combination of values of n and N-n agree with those of Hahn [7], computed using the multivariate t-distribution, our computed values agree with his to the number of significant figures given.

The acceptable quality level percentile AQLP defined as $z_p = -(\mu-L)/\sigma$, with p equal to AQL, the acceptable quality level, and related to z_p through $p = \Phi(z_p)$, was determined iteratively by requiring equality in (2.5) with $\underline{P}$ given by (4.3). Muller's [10] method, discussed above, was used in this iterative determination.

REFERENCES

[1] Cooper, B.E., "The Integral of a Non-Central t-Distribution," Journal of the Royal Statistical Society, Ser. C. 17, No. 2 (1968), 193-4.

[2] Ellis, E.W., "Double Zero Attribute Sampling Plans," Annual Technical Conference Transactions, American Society for Quality Control (1966), 340-7.

[3] Fertig, K.W. and Mann, N.R., "Finite-Lots Variables Sampling Plans Applicable When the Underlying Manufacturing Process is Normal," Research Report RR72-03, Rocketdyne, Canoga Park, Calif., 1972 (submitted for publication).

[4] Fertig, K.W. and Mann, N.R., "A Decision-Theoretic Approach to Defining Variables Sampling Plans for Finite Lots: Single Sampling for Exponential and Gaussian Processes," Journal of the American Statistical Association, 69 (September 1974), 665-71.

[5] Fertig, K.W. and Mann, N.R., "A Prediction-Interval Approach to Defining Variables Sampling Plans for Finite Lots Required to be of High Quality: Single Sampling for Gaussian Processes," Research Report RR74-07, Rocketdyne, Canoga Park, Calif. 1974.

[6] Fertig, K.W. and Mann, N.R., "The Determination of One-Sided Prediction Intervals for Normal and Lognormal Distributions, with Tables," in Reliability and Fault-Tree Analysis, Eds. R. Barlow and P. Chatterjee, SIAM Series in Applied Mathematics, 1975.

[7] Hahn, G.J., "Factors for Calculating Two-Sided Prediction Intervals for Samples from a Normal Distribution," Journal of the American Statistical Association, 64 (September 1969), 878-88.

[8] Lieberman, G.J. and Owen, D.B., Tables of the Hypergeometric Probability Distribution; Stanford University Press, 1961.

[9] Mann, N.R., Schafer, R.E. and Singpurwalla, N.D., Methods for Statistical Analysis of Reliability and Life Data: John Wiley, 1974.

[10] Muller, D.E., "A Method for Solving Algebraic Equations Using an Automatic Computer," Mathematical Tables and Computations, 10 (1956), 208-15.